Spatial Maps of Prostate Cancer Transcriptomes Reveal an Unexplored Landscape of Heterogeneity

Emelie Berglund[1,§], Jonas Maaskola[1,§], Niklas Schultz[2,§], Stefanie Friedrich[3,§], Maja Marklund[1], Joseph Bergenstråhle[1], Firas Tarish[2], Anna Tanoglidi[4], Sanja Vickovic[1], Ludvig Larsson[1], Fredrik Salmén[1], Christoph Ogris[3], Karolina Wallenborg[2,5], Jens Lagergren[6], Patrik Ståhl[1], Erik Sonnhammer[3], Thomas Helleday[2,*] and Joakim Lundeberg[1,*]

1. Department of Gene Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, Royal Institute of Technology (KTH), Science for Life Laboratory, Solna, Sweden.
2. Department of Oncology-Pathology, Karolinska Institutet (KI), Science for Life Laboratory, Solna, Sweden.
3. Department of Biochemistry and Biophysics, Stockholm University, Science for Life Laboratory, Solna, Sweden.
4. Department of Clinical Pathology, University Uppsala Hospital, Uppsala, Sweden.
5. Eukaryotic Single Cell Genomics, Science for Life Laboratory, Solna, Sweden.
6. Department of Computational Biology, School of Computer Science and Communication, Royal Institute of Technology (KTH), Science for Life Laboratory, Solna, Sweden.

[§]These authors contributed equally to this work.

[*]Correspondence should be addressed to T.H (thomas.helleday@scilifelab.se) or J.L. (joakim.lundeberg@scilifelab.se)

Abstract

Intra-tumor heterogeneity is one of the biggest challenges in cancer treatment today. Here we investigate tissue-wide gene expression heterogeneity throughout a multifocal prostate cancer using the spatial transcriptomics (ST) technology, which quantifies an array of transcriptomes across a tissue section. Utilizing a novel approach for deconvolution, we analyze the transcriptomes of nearly 6750 tissue regions and extract distinct expression profiles for the different tissue components, such as stroma, normal and PIN glands, immune cells and cancer. We distinguish healthy and diseased areas and thereby provide insight into gene expression changes during the progression of prostate cancer. Compared to pathologist annotations, we delineate the extent of cancer foci more accurately, interestingly without link to histological changes. Furthermore, we identify gene expression gradients in stroma adjacent to tumor regions that allow for re-stratification of the tumor microenvironment and increase the understanding of the development of individual prostate cancer loci within the tissue. The establishment of these profiles is the first step towards an unbiased view of prostate cancer and its microenvironment and can serve as a dictionary for future studies.

Prostate cancer (PCa) is the most common type of cancer among men worldwide[1], causes annually over 250,000 deaths, and advances via clonal evolution[2]. Frequently, multiple competing clones with independent tumor origins exist within the primary tumor[3–5]. Moreover, acquired somatic events by the clones increase the probability for metastasis[6]. Consequently, PCa contains substantial intratumoral heterogeneity with genetic alterations present both in the original tumor and in distant metastases[7,8]. Subclonal diversity can be analyzed with DNA-sequencing of a bulk tumor sample[7,9] or more precisely by using laser capture technology[3]. While genetic changes are important to track cancer heterogeneity and clonal evolution they may also be of clinical relevance. This is exemplified by the high proportion of castration-resistant PCa cases being DNA repair deficient[10]. In these cases frequent mutations in BRCA2 and ATM genes are linked to successful treatment with PARP inhibitors[11]. Furthermore the tumor microenvironment in the form of reactive stroma plays a functional role during inflammation and in neoplastic transformation[12].

By using single-cell RNA-Seq (scRNA-Seq), intra-tumor gene expression heterogeneity has been documented at the level of individual cells[13–15] and advancements in droplet microfluidics and barcoding have made it possible to analyze thousands of cells[16]. The lack of spatial information for scRNA-Seq data can, to a certain extent, be circumvented by computational inference[17,18]. Current in situ sequencing techniques have, until recently, been limited to measuring small numbers of genes[19,20] and the spatial dimensions of entire transcriptomes remain unexplored in PCa, along with the tumor microenvironment.

Pathological severity of prostate adenocarcinoma, despite progress with molecular markers and MRI, is generally scored according to the Gleason grading (Gs) system, which uses histological data only, frequently complemented with PSA measurement in blood and tumor staging[21]. However, this classification method has limitations and new alternatives have been proposed[22].

Here we investigate for the first time a multifocal PCa simultaneously at tissue- and transcriptome-wide scale using the recently introduced Spatial Transcriptomics (ST) method[23], which allows for quantification of the mRNA population in the spatial context of intact tissue. We use a novel computational procedure to elicit spatial, transcriptome-wide expression patterns enabling deconvolution of molecular events in cancer and associated microenvironment.

Results

Measuring spatial gene expression in prostate cancer tissue sections

The study design involves twelve spatially separated biopsies taken from a cancerous prostate after radical prostatectomy (Gs 3+4, pT3b, PSA = 7.1)(Fig. 1a). We measured spatial gene expression throughout twelve tissue sections using the ST methodology (Fig. 1b, Supplementary Fig. 1a). Supplementary Table 1 contains a summary of the data evaluation. Overall, 5 910 tissue regions within the 12 sections were analyzed.

Transcriptome heterogeneity in the spatial vicinity of a cancer

We initially analyzed one tissue section containing a tumor (sample 1.2, Gs 3+3, Supplementary Fig. 1b). We used a novel factor analysis method (Supplementary methods) to infer activity maps (Fig. 2a, Supplementary Fig. 2a, Supplementary data 1) and expression profiles (Supplementary Fig. 3, Supplementary data 1). The factors' activity maps generally exhibit spatial patterns that closely mirror histologically identifiable structures, such as normal glands and stroma (Fig. 2a, Supplementary Fig. 2a). Others overlap with regions annotated as cancerous or prostatic intraepithelial neoplasia (PIN) (Fig. 2a,b). Remaining factors are annotated by combined analysis of listed top genes, histology and calculated proportion of stroma, epithelium and lumen (Supplementary Fig. 2a, Supplementary Table 7). Notably, the "cancer" factor is active in a region that encompasses the annotated cancer region. Hierarchical clustering of the ST read count data (Supplementary Fig. 4a) is consistent with the factor activity patterns (Fig. 2a). In addition, principal component analysis (PCA) confirms clear separation between regions (Fig. 2d). The gene expression profiles (Supplementary Fig. 3) generally reflect the expression phenotype of the prostate as tissue of

origin or the functional requirements of the respective tissue components. For example, KLK3, KLK2, MSMB, and ACPP are among the highest-expressed genes in many of the factors. In the "stromal" factor the top-expressed genes have functions related to cytoskeleton, smooth muscle and cell-adhesion. Known PCa-related genes (SPON2, TFF3, SPINK1) are among the highest-expressed genes in the "cancer" factor.

In order to identify interactions between the factors, we performed hierarchical clustering (Supplementary Fig. 2b) of all ten factors. The clustering yielded three main groups. Among the ten factors, inflammation, PIN and cancer, could delineate one of the groups, while being seperated from normal glands. This further indicate inflammation as a critical component of tumour initiation and progression. As expected, factors containing mainly stroma cells cluster together. Factor 1 contained high levels of MSMB and annotated as "normal glands signature". MSMB is known to be downregulated in prostate cancer[24].

The gene expression within each region (normal, cancer, PIN) obtained from the preceding factor analysis was investigated to identify region-specific markers (Supplementary Fig. 5b). In the cancer region we for instance observe enrichment of SPINK1 and PGC, and depletion of ACPP (Fig. 2c, Supplementary Table 2). Another noteworthy observation is elevated levels of NPY in the PIN region (Fig. 2c). To validate these findings we investigated the concordance between gene expression and staining of the corresponding proteins within the tissue. Immunostaining of SPINK1 was coincident with the defined cancer region and immunostaining of NPY was mostly localized to that of the PIN region (Fig. 2e,f).

Spatial expression patterns common to cancer tissue sections

Next, we carried out a factor analysis for three tissue sections containing annotated cancer foci (Fig. 3). The resulting gene expression profiles (Supplementary Fig. 5, Supplementary Data 2) are similar to the preceding analysis. The activity maps (Fig. 3b, Supplementary Fig. 6a, Supplementary Data 2) again show patterns corresponding to annotated or histologically identifiable structures (Fig. 3a), but now group regions with similar phenotype across tissue sections, and for tissue section 1.2 are virtually identical to the preceding analysis. In particular, the "cancer" factor now includes in section 2.4 a region annotated as suspected cancer and also shows slight activity in section 3.3 (Fig. 3b). SPON2, TFF3 and SPINK1 are again among the highest-expressed genes for this factor (Supplementary Fig. 5). The gene expression profile of another factor reflects processes observed in reactive stroma ("reactive stroma", Supplementary Fig. 5) and surrounds that of the "cancer" factor in samples 2.4 and 3.3 (Fig. 3b). Although the pathologist marked section 3.3 with cancer and few PIN and normal glands, we detected relatively large areas containing both normal and PIN glands within the cancer annotated region ("normal glands" and "PIN glands", Fig. 3b). This heterogeneity within the cancer area is supported by PCA and hierarchical clustering of spots taken from the annotated cancer area in sample 3.3 (Supplementary Fig. 7).

Notably, hierarchical clustering (Supplementary Fig. 6b) of the factors was similar to the preceding one (Supplementary Fig. 2b). Importantly, inflammation is once again linked to cancer and PIN, and a clear separation of normal glands (with and without MSMB) from stroma and malignant cells was observed. Interestingly, we identified factor 1 to be a mix of stroma and normal epithelial cells expressing high levels of RACK1 and FTH1. Studies has

shown that RACK1 is part of the tyrosine kinase signaling, facilitating transcriptional activity of the androgen receptor (AR)[33].

In addition to our genome-wide measurements, we investigated pathways in normal, PIN and cancer epithelium (Fig. 3c, Supplementary Fig. 8, Supplementary Table 4). Higher expression of cancer-related pathways was observed in cancer and PIN compared with normal epithelium. We observed increased expression of cell cycle, DNA replication, homologous recombination, Fanconi anemia and p53 signaling in the cancer and PIN areas. Notably, Wnt signaling was only detected in the PIN factor. This suggest that Wnt signaling plays an important role in the initiation and progression of PIN. Studies have shown association of Wnt signaling/β-catenin expression with tumor evolution[34]. All of the above mentioned pathways differentiate benign from malignant tissue and with this in mind, they could be particularly important in characterizing normal or pre-neoplastic tissue (PIN). It also opens up the possibility to develop anticancer treatments to target these pathways, for example cell cycle arrest that hinder progression of cells from the G_1 to S phase[35,36] or targeting DNA repair with PARP inhibitors[11]. Moreover, tumours with deficiencies in homologous recombination or Fanconi anaemia proteins are known to respond to platinum-based chemotherapy[37].

Disappearance of basal cells from prostatic glands is one of the hallmarks of PCa, and the protein P63 is a well-established marker for basal cells[38]. We validated by IHC that presence or absence of P63-stained glands co-incident with pathological annotation and the ST factor analysis results (Supplementary Figs. 9-12). Notably, the absence of P63 staining in sample

1.2 is coincident with the "cancer" factor's activity, and extends further than indicated by the pathologist (Supplementary Fig. 10). In sample 2.4, no P63-stained glands are observed throughout an area that comprises the annotated and suspected cancer foci (Supplementary Fig. 11). While P63 staining is mostly absent in the large annotated cancer area of sample 3.3, some stray glands show P63 staining (Supplementary Fig. 12). Closer inspection reveals flat glands with P63 staining (Supplementary Figs. 9 and 12), consistent with PIN characteristics[39]. To confirm the ST expression data for several genes of interest, we performed immunohistochemistry (IHC) and found concordance between the spatial expression patterns of protein and mRNA (Supplementary Figs. 13-14).

To explore observation across tumors, we analyzed two more tumors from two patients (corresponding to an additional 750 tissue regions) and compared the transcriptional profiles of the tumor areas from the first patient (Supplementary Fig. 15a, Supplementary Data 7). We detect extensive tumor heterogeneity between patients as determined by the factor analysis. To further investigate, at the gene level, we could confirm some observations obtained from patient 1. Interestingly, we identified NPY (enriched in PIN regions), SPON2 (enriched in cancer regions) and NR4A1 (enriched in reactive stroma, see below) also in patient 2 and 3 (Supplementary Fig. 15b), while EEF2, NEAT1 and TPT1 (established interaction with p53)[40] were uniquely expressed in patient 3 (Supplementary Fig. 16).

The transcriptome wide data for patient 2 looked quite distinct compared to patient 1 and 3. Among the highest genes, we found SERPINA3 and TPT1 (Supplementary Fig. 16).

Gene expression differences in the center and the periphery of cancer

We sought to in more detail investigate functional differences in gene expression between the center and the periphery of the cancer, respectively and how the signals from the tumor stimulates the adjacent endothelium. Abnormal tissues adjacent to tumors were first described in 1953, also called the "field cancerization" (Slaughter et al., 1953). Previous studies have suggested that breast cancer tissue close to cancer undergoes extracellular matrix remodeling, fibrosis, and an epithelial-to-mesenchymal transition (EMT) (Trujillo et al., 2011). Other studies, focusing on prostate cancer investigated the gene expression differences among prostate cancer tissue, adjacent prostate cancer tissue and normal prostate tissue (Chandran et al., 2005). They found that the tumor vs normal expression profile was more extensive than the tumor vs adjacent normal profile. Also, tumor and adjacent tumor tissues emerged with higher response of inflammatory and immune than normal tissues, which was in agreement with previous report that inflammation was closely related to cancer. A recent study conducted on many different cancer types revealed that the adjacent tumor tissue represents an intermediate state between normal and cancer tissue (Aran et al., 2017). They also uncovered activation of pro-inflammatory response in the adjacent tissue. However, no evaluation with spatial resolution of cancer and adjacent cancer tissue has been performed to date. We therefore aimed for discover differences between the cancer and the periphery of the cancer samples 1.2 (Gs 3+3), 2.4 (suspected cancer) and 3.3 (Gs 3+4) (Fig. 4 and Supplementary Figs. 18-19).

The pathways activated in the center of sample 1.2 (Fig. 4d) are mainly linked to altered cellular metabolism (Oxidative phosphorylation, Pentose phosphate pathway[28], Citrate cycle); metabolic alteration is a hallmark of cancer[29]. The activated TCA cycle pathway is essential for a neoplastic prostate cell to evolve into a malignant tumor cell[30] and its activations in the

center suggests that the malignant center cells with high energy consumption are surrounded by pre-malignant cells. On the other hand, for sample 3.3 (Fig. 4h), higher levels of metabolism is seen in the periphery of the cancer. Furthermore, sample 3.3 expresses higher levels of Endocytosis, Phagosome and Lysosome in the central area. These pathways are known to be high in necrotic areas, in which they clear cell debris and dead cells (Green and Martinez, 2016). Notably, we also observe enrichment of the HIF-1 signaling pathway. As a tumor grows, it causes abnormalities in tumour blood vessels, leaving region of the tumor with lower oxygen concentration compared to normal tissue (Vaupel et al., 2009). The best understood mechanism of how cancer cells adapt to a hypoxic environment is through elevated levels of HIF-1 and HIF-2 (Semenza et al., 2014).

Notably, the activated pathways in the periphery of 1.2 are mainly related to stress, inflammation (NF-kappa B signaling, Toll-like receptor signaling, NOD-like receptor signaling) (Fig. 4d). Such theory is supported by other studies (Chandran et al., 2005 and Aran et al., 2017). We confirmed that the presence of immune pathways was due to presence of immune cells in the nearby tissue by asking a pathologist to annotate if inflammation exists (Supplementary Fig. 17a). We also found high levels of immune-related genes in the periphery (e.g. IRF7, HLA-C and NFKBIA) (Supplementary Fig. 17a).

Both tumors showed high levels of pathways linked to cell proliferation (MAPK signaling, ECM-receptor interaction, PI3K-Akt signaling) and cell motility (regulation of the actin cytoskeleton, focal adhesion), although they were found in different regions of 1.2 and 3.3 (either the center or periphery of the cancer) (Fig. 4d,h).

We see higher expression of some genes in the periphery compared to the center in all three cancer areas (Fig. 4 and Supplementary Figs. 18-19). For example, TAGLN (tumour

suppressor) (Prasad et al., 2010), HLA (linked to the immune system in humans) (Choo et al., 2007), ACTG2 and ACTB (involved in cell motility) (Consortium, Uniprot). Genes that are higher expressed in the center in all three cancer areas are for example, NUPR1, ASAH1 PDLIM5, KLK4 and PSCA. All are known to be highly expressed in PCa (Uhlen et al., 2017). Additionally noteworthy are EGR1 and KLK3 in sample 3.3 (Gs 3+4). EGR1 is up-regulated in the center of the cancer area of sample 3.3 and there is evidence suggesting that this gene is directly linked to the transition of the cancer into invasive carcinoma and a potential target for cancer treatment (Abdulkadir et al., 2001, Baron et al., 2003). KLK3 is to date the best prostate cancer biomarker. However, its downregulation as observed in sample 3.3 (Gs 3+4) is linked to more aggressive tumours and recent studies have shown that the stimulation of KLK3 expression can indirectly reduce the proliferation rate and decrease the risk of metastasis (Diamandis et al., 2000).

The cancer area of sample 2.4 shows similarities with the one in sample 1.2 (Fig. 4 and Supplementary Fig. 18). The center of 2.4 is dominated by enriched pathways linked to altered cellular metabolism (Glycolysis / Gluconeogenesis, Pyruvate metabolism, Amino sugar and nucleotide sugar metabolism)[29], whereas the periphery is marked by enriched pathways involved in stress, inflammation and immune system (B cell receptor signaling, T cell receptor signaling, Toll−like receptor signaling) (Chandran et al., 2005 and Aran et al., 2017). Consistent with the factor intensities, the cancer area of sample 2.4 shows more pathway similarities with sample 1.2 than with sample 3.3 (Fig. 4 and Supplementary Fig. 18). The pathway Protein processing in endoplasmic reticulum is the most significantly enriched pathway in all three cancer centers. Due to increased cellular metabolism and cell proliferation rate, cancer cells activate this pathway to obtain correct protein synthesis and modifications and overcome ER stress (Yadav et al., 2014).

Spatial expression patterns in the microenvironment of cancer and inflammation

To examine whether we could identify inflammation, we performed analysis of all 12 samples. The data revealed a factor ("inflammation" factor, Supplementary Fig. 19, Supplementary data 3) that was active across inflammation-annotated regions of sample 3.1, and 4.2. A more granular factor analysis of only these two samples decomposes several stromal and glandular factors (Fig. 5, Supplementary Fig. 20, Supplementary Data 4). We detect spatial proximity of areas consisting of apparent cancer glands ("FOSB-enriched"), inflammation ("AQP3-enriched") and reactive stroma ("NR4A1-enriched") (Fig. 5b). A novel gene (NR4A1) was highly expressed in reactive stroma. We confirmed the spatial expression pattern of NR4A1 by its protein by IHC (Supplementary Fig. 21). Among the factors not shown in Fig. 5, some contained stroma cells or normal glands (with or without MSMB) (Supplementary Fig. 22). KRT13 was upregulated in one of the factors and has recently been associated with poor prognosis in metastatic patients[41].

Interestingly, both samples exposes stromal expression gradients adjacent to tumor regions (Fig. 5c) which are likely due to stromal cell-type heterogeneity. For the inflammation containing samples, hierarchical clustering of the factors revealed three distinct groups (Supplementary Fig. 22b). Stroma and normal cells separate from inflammation, however, compared to the previous clustering (Supplementary Fig. 2b, Supplementary Fig. 6b) reactive stroma is more similar to inflammation and cancer glands ("FOSB enriched") than stroma cells.

Whereas molecular alterations drive the progression from low-Gleason grade to invasive cancer, tumor microenvironment and tumor cells are co-dependent and progress alongside.

Hence, we compared both the gene expression on stroma close to tumor and inflammation with that of normal stroma (Fig. 5d, Supplementary Tables 5-6). The results display different enriched pathways between the two regions. Normal stroma was associated with cell movement and adhesion (actin cytoskeleton and regulation of Actin-based Motility by Rho) as well as androgen signaling and the complement system. Several studies have suggested that the complement system is involved in the immuno-surveillance against tumours (anti-tumor effect)[42]. On the other hand, the complement has also been implicated in tumor growth[43]. Our results elucidated that the complement pathway is enriched in stroma cells close to normal epithelium. The reactive stroma was enriched for oxidative stress and ILK signaling. Studies have shown that ILK expression and activity is significantly up-regulated in several types of cancers (pro-tumor effect)[44].

To further study cancer and inflammatory microenvironments, we selected samples based on activity of the "reactive stroma" factor in the 12 sample analysis (Supplementary Fig. 19), Supplementary Data 3). A factor analysis of the four identified samples, and subsequent dimensionality reduction, reveal once again cancer, reactive stroma and inflamed glands in close proximity to each other (Supplementary Figs. 23-25, Supplementary Data 5).

In order to enable other reserachers to study their gene of interest on the tissue, a Shiny application was built and is freely available at https://spatialtranscriptomics3d.shinyapps.io/STProstateResearch/ (Supplementary Methods).

Deletions and amplification are spread locally

In order to complement the spatial gene expression patterns, we examined the copy number from whole genome sequencing data (Supplementary Fig. 26). We analysed the affected base pairs in exonic regions per copy number values below two for each sample to gain insights about the genetic structure of each sample and to link deletions to cancerous areas. Sample 3.3 (Gs 3+4) shows more deleted base pairs above a CNV of 0.8 (15.6 kbp affected). We conclude that up to 50% of the cells in the tissue are affected by deletions if homozygous deletions are assumed. For sample 2.4 (Gs 3+4 & suspected Gs), higher number of deleted base pairs occurs with copy number of 1.3 (18 kbp affected) which appropriates 25% of the cells if only homozygous deleted segments are considered. The number of cells with deletions in the samples 2.3 and 3.1 is 35% and 25% respectively. Sample 1.2 (Gs 3+3) shows no higher number of deleted base pairs. None of the samples considered as histological normal show an increased occurrence of deleted segments. The samples that show an increased occurrence of deleted base pairs in exonic regions per copy numbers contain either cancerous or larger inflamed areas.

A similarity tree based on Euclidean distance and hierarchical clustering was created (Supplementary Fig. 27). The tree reflects the relation of deleted and amplified segments within the whole genome of the twelve tissue sections. Four clusters were revealed and each cluster contains one cancerous sample. We conclude that the genetic structural variations appear unique for each cancer sample. Sample 3.3 shows the largest difference to germline and contains the highest number of genetic structural variations of the twelve samples. Interestingly, we observe that the four cancer samples do not share many deletions or

amplifications pointing to independent tumor origins. We compared the samples and their assigned clusters with their physical position within the prostate (Supplementary Fig. 27). For three clusters we observe that the samples, which belong to one cluster, are physically close to each other. It can be concluded that the deletions and amplifications are mostly spread locally.

Finally, we sought to assess the relationship of copy number and gene expression. Concerning amplifications and deletions, each cancerous or inflamed tissue sample shows a unique genetic structure (Fig. 6). In general, sample 1.2 is rather shaped by expressed genes with amplifications, sample 3.3 by deletions, and sample 2.4 harbors both, deletions and amplifications. Further, genes with amplifications or deletions are expressed primarily in small regions within the tissue mirroring a genetically heterogeneous tissue.

Discussion

Here we investigate tissue-wide gene expression heterogeneity throughout a multifocal PCa using the ST technology which quantify an array of transcriptomes across a tissue section. To ensure that patient genotype do not confound the analysis, we have selected a single prostate that is analysed in a comprehensive manner, analyzing more than 6 000 tissue regions. Compared to the spot diameter, the lateral RNA diffusion is negligible under the tissue slide, ensuring that the measured gene expression stems from the local tissue. The number of mRNA molecules obtained is in line with previous reports[23], and on average we sequence 2 million unique reads on a tissue slide. Hence, we measure, for the first time, spatial gene expression in PCa tissue sections. Further evidence for the accuracy of the spatial gene

expression measurements is presented by concordance between gene expression and staining of the corresponding protein within the tissue.

We have developed an unsupervised probabilistic framework to analyze spatial transcriptomics data. The strength of this method lies in that the method takes the spatial position into account, which is lost in other used methods such as PCA and t-SNE. We identify factors corresponding to the different parts of tissue architecture, such as cancer, inflammation, normal or reactive stroma and normal or PIN glands, etc. Interestingly, we observe that distinct cancer expression regions can extend beyond the boundaries of annotated tumor areas. Similarly, the discovered gene expression profiles may be used to predict further regions of potential cancer, PIN or inflammation. From a clinical perspective, we suggest that our analysis may be used to alert pathologists to give extra attention to 'high risk' areas based on localized, transcriptome-wide data. Furthermore, we observe clear separation of gene expression patterns between normal prostate epithelium and cancer areas with elevated Gs (3+3 and 3+4), as expected[45], suggesting that this approach may be useful to gain further understanding of human tumor in situ. Importantly, we report on heterogeneity within one patient as well as heterogeneity between patients. Thus, this study highlights the value of focusing on individual profiling.

The current investigation also provides new insights into gene expression differences between cancer core and periphery and pose significant questions that have important implications for the development of prostate cancer. A related question is whether the periphery of the tumor promotes tumor initiation and progression and if so may provide new tools for early detection.

Finally, we argue that a new landscape is revealed by spatially mapping gene expression and analyzing it in an unsupervised manner. For instance, it enables de-novo characterization and delineation of reactive stroma in the proximity of cancer and inflammation (Fig. 5) to elucidate the role of the microenvironment for PCa. Our study extends previous suggestions that inflammation and reactive stroma are evident at the earliest stages of neoplastic progression, stimulating development of the tumor[46] to demonstrate spatially confined gene expression gradients. Indeed, changes in the microenvironment may even precede genetic alteration in the tumor core. This study uncovers high levels of oxidative stress and ILK signaling within the reactive stroma whereas normal stroma was associated with cell movement and adhesion (actin cytoskeleton). High levels of oxidative stress in reactive stroma indicate that the cancer is dependent upon that the stroma releases energy to fuel cancer cells and enable growth and survival. Our spatial analysis describes reactive stroma as an emerging hallmark of cancer initiation and progression.

Taken together, our results have revealed that an analysis of tumor gene expression in a spatial context dramatically increases the granularity compared to a bulk analysis. Sampling different parts of the same tumour showed remarkable differences on the transcriptome level of the cancer cells at each site. Thus this massive tissue region analysis could serve as a foundation for a transcriptome-based clinical evaluation of cancer tissues as well as provide insight into gene expression in the tumor microenvironment. In summary, we propose that expression profiles based on spatial factor analysis can serve as a tissue-transcriptome dictionary as basis for a wide range of new scientific discoveries in prostate cancer.

Methods

Collection and preparation of prostate cancer tissue

Radical prostatectomy was performed on a patient with adenocarcinoma (Gs 3+4, pT3b, PSA 7.1). The whole prostate was snap-frozen in liquid nitrogen and twelve sections covering half of the prostate were cryosectioned at 10 µm thickness. Sections were placed on prepared glass slides and incubated at 37 ℃ for 1 minute, fixated in 16% methanol-free formaldehyde (#8906, Thermo Fisher Scientific) and washed in PBS. We applied the current ST protocol to all sections, yielding both traditional H&E images amenable to annotation by pathologists, as well as gene expression profiles for every microarray spot[23]. Tissue sections were annotated for pathological status and four were judged to exhibit tumor foci. Three of the tumor foci are situated in tissue regions covered by array spots.

Staining and imaging

These steps were described previously[23]. Tissue was incubated in Mayer's Hematoxylin for 5 minutes and in eosin (1:20 in Tris buffer, pH 6) for 20 s.

Quality control array experiments and detection of fluorescent cDNA "footprint"

Prior to spatial barcoding experiments, permeabilization conditions were optimized in a quality control experiment for maximal mRNA yield in PCa tissue (Supplementary Fig. 28). The fluorescent cDNA was consistent with the tissue structure shown by histology.

Quality control experiments were carried out as described in section permeabilization and reverse transcription, except for some minor changes. This test was performed in order to

study the optimal permeabilization conditions. The array was printed to have all sequences necessary for capture of mRNA but without a spatial barcode. The release step of the surface probes was not performed and the reverse transcription mixture contained the same reagents except for 0.5 mM of each dATP/dGTP/dTTP, 12.5 µM dCTP and 25 µM Cy3-dCTP. The procedure involved testing times of 8, 10 and 12 minutes. Glass slides were scanned in an InnoScan 910 scanner (Innopsys) with 5 µm resolution and gain at 10%. Signal intensities were investigated using the Mapix software.

Permeabilization and reverse transcription

These steps were described previously[23]. Exonuclease pre-permeabilization was performed at 37 ℃ for 30 minutes, followed by an incubation in pepsin at 37 ℃ for 10 minutes. Reverse transcription was performed at 42 ℃ overnight. In order to remove and degrade the prostate tissue, a mix of 1% β-mercaptoethanol in RLT buffer was added to the samples and incubated at 56 ℃ for 1 hour and 15 minutes. A second removal mix containing Proteinase K in PDK Buffer was added to each well and incubated at 56 ℃ for 1 hour and 15 minutes. The release step was done at 37 ℃ for 1 hour and 15 minutes. Once the surface probes were de-attached, 65 µl from each well was collected. After probe release, the features with non-released DNA oligonucleotide fragments were detected by hybridization and imaging, as previously described[23], in order to obtain Cy3-images for alignment. The bright field images and fluorescent images were manually aligned using Adobe Photoshop CS6 (Adobe) by utilizing the visible spots and structures from both images.

Library preparation of cDNA for sequencing

The steps were performed as earlier described[23]. Finished libraries were diluted to 4 nM and sequenced on the Illumina HiSeq or NextSeq platform using paired-end sequencing. Typically, 31 or 101 bases were sequenced on read one to determine the spatial barcode, and 121 or 101 bases were sequenced on read two to cover the genetic region. ST sequencing reads were mapped against the human genome (GRCh38), and Ensembl (release 85) transcripts were quantified, as described previously[47].

Dissecting spatial gene expression patterns with factor analysis

We have developed a core model to perform factor analysis applicable for spatial gene expression data. The full model is described in Supplementary methods and Supplementary Data 6. This method (unsupervised) needs no prior knowledge of reference expression data. It seeks to factor the gene expression into spatial factor activity maps and gene expression profiles. The factor activity maps reflect the amount of mRNA contributed by a given factor in every spot and are useful for visual inspection and comparison to morphological features. The expression profiles quantify how strongly each gene is expressed in a given factor and are thus informative about biological processes. The discovered "factors" correspond to cell types, but aside from biological effects, the analysis method also captures technical effects (such as sequencing depth of the libraries) and can allow for correction of such unavoidable artifacts. This enables us to quantify gene-expression differences between individuals, genotypes, developmental stages and disease states. Throughout the manuscript we name factors according to histological features co-incident with a factor's spatial activity or according to highly- and specifically-expressed genes.

The factor analysis software is available under the GNU General Public License v3 at https://github.com/maaskola/spatial-transcriptome-deconvolution.

Data availability

Sequencing data have been deposited on dbGaP (https://www.ncbi.nlm.nih.gov/gap) under the accession code XX (embargoed until publication). Count matrixes are available at http://www. spatialtranscriptomicsresearch.org/.

Acknowledgments

This study was supported by AstraZeneca and Science for Life Laboratory. We thank National Genomics Infrastructure (NGI), Sweden for providing infrastructure support. The data were analyzed using resources provided by SNIC through the Uppsala Multidisciplinary Center for Advanced Computational Science (SNIC/UPPMAX).

We thank Annelie Mollbrink for providing lab assistance, Mobashir Mohammad and Sailendra Pradhananga for performing inital analyses, and Johan Lindberg, Pelin Sahlén and Carsten Daub for comments on the manuscript.

Author information

Contributions

Experiments were conceived and designed by E.B.. Experiments were performed by E.B., N.S., K.W. and M.M.. Computational methods were developed by J.M., with input from J.Lagergren. Bioinformatics analyses were performed by E.B., J.M., S.F., C.O.,

L.L., F.S., S.V., and J.B.. A.T. inspected the tumor samples and performed morphological annotations. F.T. provided samples and biological input. The manuscript was written by E.B., J.M., T.H., S.F., E.S. and J.L. with input from all authors.

Competing interests

P.L.S., F.S., and J.L. are authors on patents applied for by Spatial Transcriptomics AB covering the technology.

Corresponding author

Correspondence to J.Lundeberg.

Electronic supplementary material

Supplementary Information

Supplementary Description

Supplementary Data 1

Supplementary Data 2

Supplementary Data 3

Supplementary Data 4

Supplementary Data 5

Supplementary Data 6

Supplementary Data 7

Supplementary Table 1-6

References

1. Bray, F., Ren, J.-S., Masuyer, E. & Ferlay, J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* 132, 1133–1145 (2013).

2. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366, 883–892 (2012).

3. Lindberg, J. *et al.* Exome Sequencing of Prostate Cancer Supports the Hypothesis of Independent Tumour Origins. *Eur. Urol.* 63, 347–353 (2013).

4. de Bruin, E. C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346, 251–256 (2014).

5. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet.* 47, 367–372 (2015).

6. Gupta, G. P. & Massagué, J. Cancer metastasis: building a framework. *Cell* 127, 679–695 (2006).

7. Hong, M. K. H. *et al.* Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nature Communications* 6, (2015).

8. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* 520, 353–357 (2015).

9. Morrissy, A. S. *et al.* Spatial heterogeneity in medulloblastoma. *Nat Genet.* 49, 780–788 (2017).

10. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215–1228 (2015).

11. Mateo, J. *et al.* DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *N. Engl. J. Med.* 373, 1697–1708 (2015).

12. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674 (2011).

13. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401 (2014).

14. Kim, K.-T. *et al.* Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol.* 16, 127 (2015).

15. Grun, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255 (2015).

16. Junker, J. P. & van Oudenaarden, A. Single-cell transcriptomics enters the age of mass production. *Mol. Cell* 58, 563–564 (2015).

17. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotech.* 33, 495–502 (2015).

18. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotech.* 33, 503–509 (2015).

19. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Meth.* 10, 857–860 (2013).

20. Lee, J. H. *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* 343, 1360–1363 (2014).

21. Epstein, J. I. *et al.* The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma:Definition of Grading Patterns and Proposal for a New Grading System. *Am. J. Surg. Pathol.* 40, 244–252 (2016).

22. Gordetsky, J. & Epstein, J. Grading of prostatic adenocarcinoma: current state and prognostic implications. *Diagn. Pathol.* 11, (2016).

23. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82 (2016).

24. Whitaker, H. C. *et al.* The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. *PLoS One* 5, e13363 (2010).

25. Vestergaard, E. M. *et al.* Promoter hypomethylation and upregulation of trefoil factors in prostate cancer. *Int. J. cancer* 127, 1857–1865 (2010).

26. Perera, O. *et al.* Trefoil factor 3 (TFF3) enhances the oncogenic characteristics of prostate carcinoma cells and reduces sensitivity to ionising radiation. *Cancer Lett.* 361, 104–111 (2015).

27. Consortium, T. U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169 (2016).

28. Tsouko, E. *et al.* Regulation of the pentose phosphate pathway by an androgen receptor-mTOR-mediated mechanism and its role in prostate cancer cell growth. *Oncogenesis* 3, e103 (2014).

29. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.* 23, 27–47 (2016).

30. Costello, L. C. & Franklin, R. B. The intermediary metabolism of the prostate: a key to understanding the pathogenesis and progression of prostate malignancy. *Oncology* 59, 269–282 (2000).

31. Yoon Seon Yoo, Hye Gyeong Han, and Y. J. J. Unfolded Protein Response of the Endoplasmic Reticulum in Tumor Progression and Immunogenicitye. *Oxid. Med. Cell. Longev.* (2017).

32. Kerr, C. Huntington's disease provides cancer clues. *The Lancet. Oncology* 3, 518 (2002).

33. Kraus, S., Gioeli, D., Vomastek, T., Gordon, V. & Weber, M. J. Receptor for activated C kinase 1 (RACK1) and Src regulate the tyrosine phosphorylation and function of the androgen receptor. *Cancer Res.* 66, 11047–11054 (2006).

34. Yu, X., Wang, Y., DeGraff, D. J., Wills, M. L. & Matusik, R. J. Wnt/β-Catenin activation promotes prostate tumor progression in a mouse model. *Oncogene* 30, 1868–1879 (2011).

35. Singh, S. K., Banerjee, S., Acosta, E. P., Lillard, J. W. & Singh, R. Resveratrol induces cell cycle arrest and apoptosis with docetaxel in prostate cancer cells via a p53/ p21WAF1/CIP1 and p27KIP1 pathway. *Oncotarget* 8, 17216–17228 (2017).

36. Chu, I. M., Hengst, L. & Slingerland, J. M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat. Rev. Cancer* 8, 253–267 (2008).

37. Fong, P. C. *et al.* Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J. Clin. Oncol.* 28, 2512–2519 (2010).

38. Weinstein, M. H., Signoretti, S. & Loda, M. Diagnostic utility of immunohistochemical staining for p63, a sensitive marker of prostatic basal cells. *Mod. Pathol.* 15, 1302–1308 (2002).

39. Bostwick, D. G. & Qian, J. High-grade prostatic intraepithelial neoplasia. *Mod. Pathol.* 17, 360–379 (2004).

40. Chen, W. *et al.* Tumor protein translationally controlled 1 is a p53 target gene that promotes cell survival. *Cell Cycle* 12, 2321–2328 (2013).

41. Li, Q. *et al.* Keratin 13 expression reprograms bone and brain metastases of human prostate cancer cells. *Oncotarget* 7, 84645–84657 (2016).

42. Yamakawa, M. *et al.* Protection of thyroid cancer cells by complement-regulatory factors. *Cancer* 73, 2808–2817 (1994).

43. Markiewski, M. M. *et al.* Modulation of the antitumor immune response by complement. *Nat. Immunol.* 9, 1225–1235 (2008).

44. Persad, S. & Dedhar, S. The role of integrin-linked kinase (ILK) in cancer progression. *Cancer Metastasis Rev.* 22, 375–384 (2003).

45. Jézéquel, P. *et al.* Validation of tumor-associated macrophage ferritin light chain as a prognostic biomarker in node-negative breast cancer tumors: A multicentric 2004 national PHRC study. *Int. J. Cancer* 131, 426–437 (2012).

46. Qian, B.-Z. & Pollard, J. W. Macrophage diversity enhances tumor progression and metastasis. *Cell* 141, 39–51 (2010).

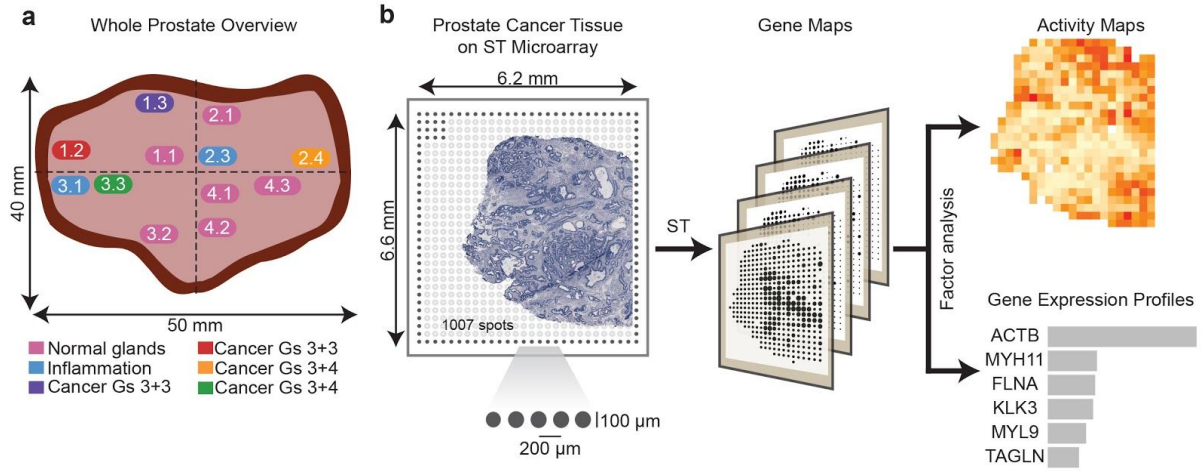
47. Fernandez Navarro, J., Sjostrand, J., Salmen, F., Lundeberg, J. & Stahl, P. L. ST Pipeline: An automated pipeline for spatial mapping of unique transcripts. *Bioinformatics* (2017).

48. SJ., C. GSTM1. *Cancer Genetics Web* (2015). Available at: http://www.cancer-genetics.org/GSTM1.htm. (Accessed: 7th December 2017)

49. Jarick, I. *et al.* Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum. Mol. Genet.* 20, 840–852 (2011).

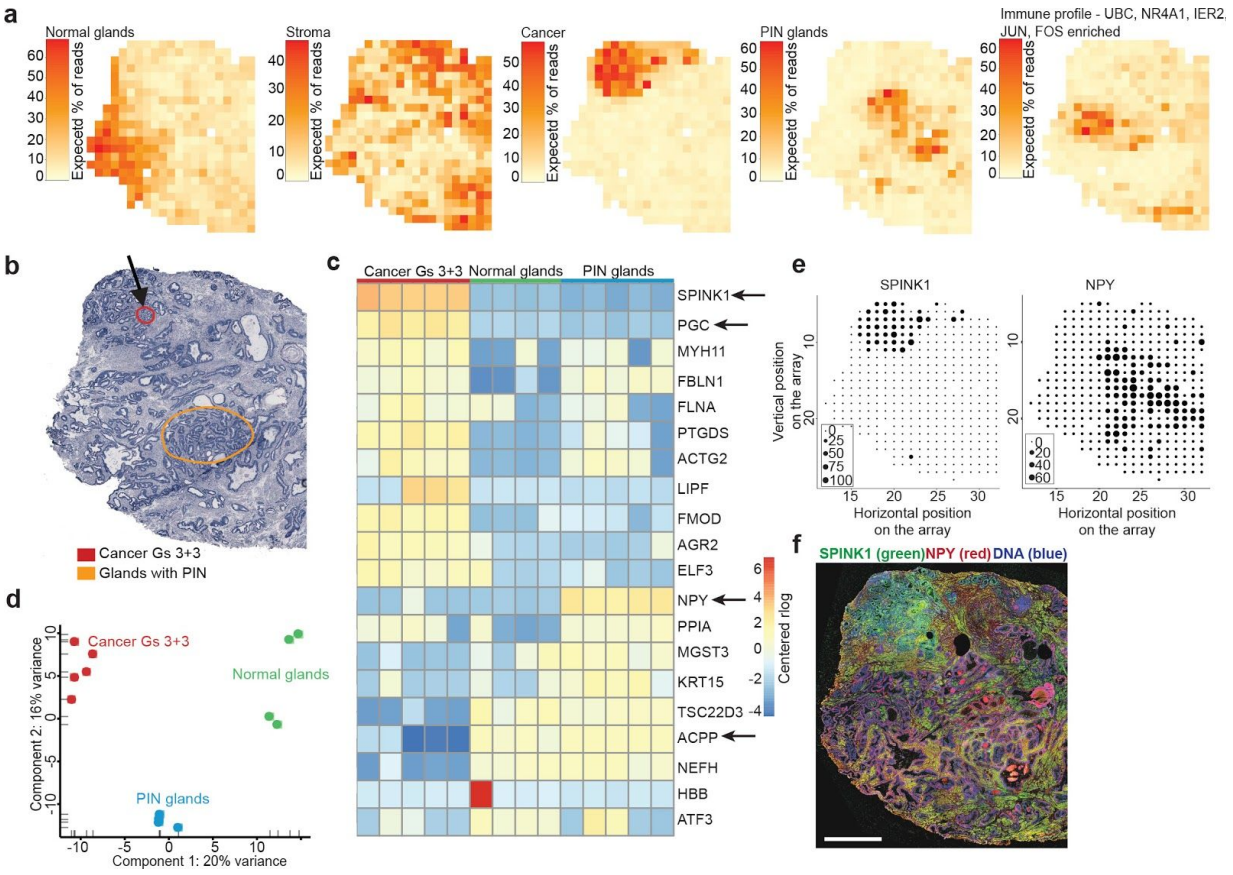
Figures

Figure 1



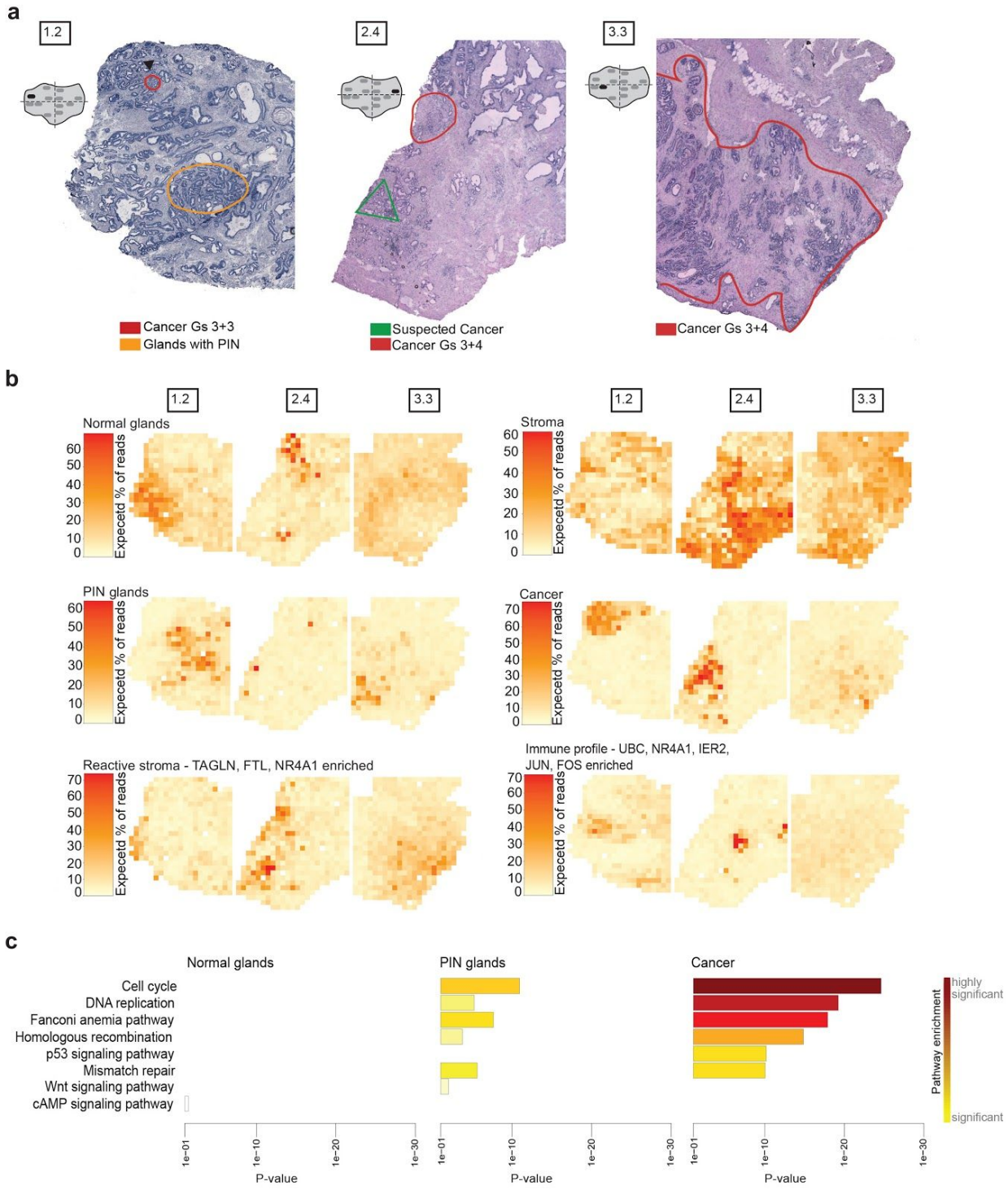
Study design for spatial transcriptomics (ST) in prostate cancer. **a** Location of sections used in this study and annotations made by a pathologist. The sections are colored according to annotation. Scale bars indicate size of the prostate. **b** Spatial microarrays have 1007 spatially-barcoded spots of 100 µm diameter and 200 µm center-to-center distance. Spots denoted by filled circles are used for orientation, and lack spatial barcodes. The ST procedure yields matrices with read counts for every gene in every spot, which are then decomposed by factor analysis resulting in a set of factors (“cell types”), each comprising one activity map and one expression profile.

Figure 2



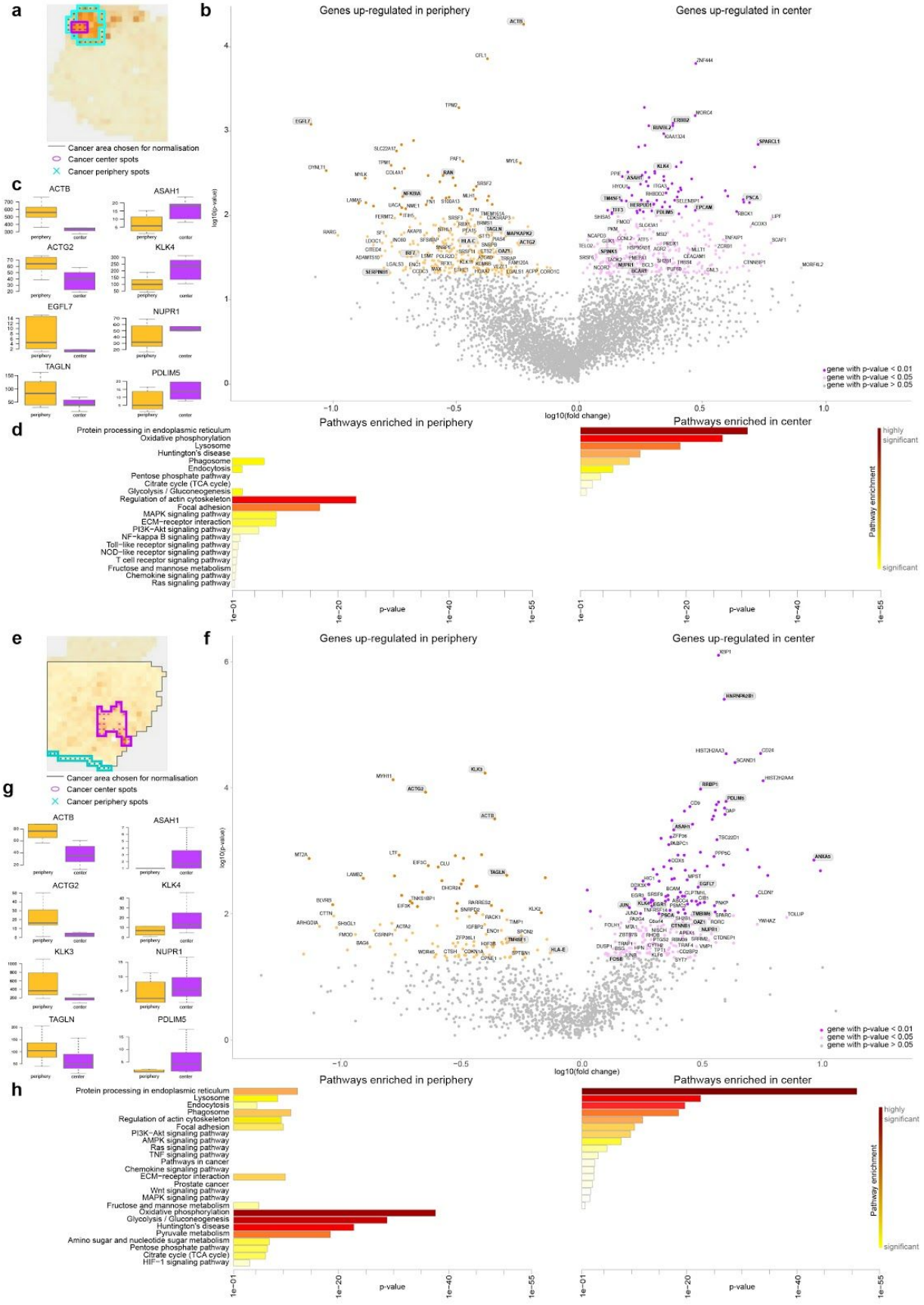
Spatial gene expression heterogeneity within the 1.2 cancer tissue sample. **a** Factor activity maps for selected factors corresponding to epithelial, stromal, cancerous, PIN, or inflamed regions. Remaining factors' activity maps in Supplementary Figure 2 and Supplementary data 1. **b** Annotated brightfield image of H&E-stained tissue section. **c** Heatmap of the 20 most variable genes between cancer, PIN and normal gland regions, using spot sets from Supplementary Fig. 4b. Centered rlog: difference of rlog (variance-stabilized transform of ST expression data) and gene-wise mean rlog. Arrows highlight genes of interest validated by immunohistochemistry (IHC). **d** First two principal components of spot sets from **c** separate cancer, PIN and normal regions. **e** Array dot plots for SPINK1 and NPY. Circle size in array dot plots indicates normalized ST counts. **f** IHC staining for SPINK1 and NPY of an adjacent section on the ST array. Nuclei are stained with DAPI (blue). Scale bar indicates 1 mm.

Figure 3


a
1.2
2.4
3.3
Cancer Gs 3+3
Glands with PIN
Suspected Cancer
Cancer Gs 3+4
Cancer Gs 3+4
b
1.2
2.4
3.3
Normal glands
Expecetd % of reads
Stroma
PIN glands
Cancer
Reactive stroma - TAGLN, FTL, NR4A1 enriched
Immune profile - UBC, NR4A1, IER2, JUN, FOS enriched
c
Normal glands
PIN glands
Cancer
Cell cycle
DNA replication
Fanconi anemia pathway
Homologous recombination
p53 signaling pathway
Mismatch repair
Wnt signaling pathway
cAMP signaling pathway
P-value
Pathway enrichment
highly significant
significant

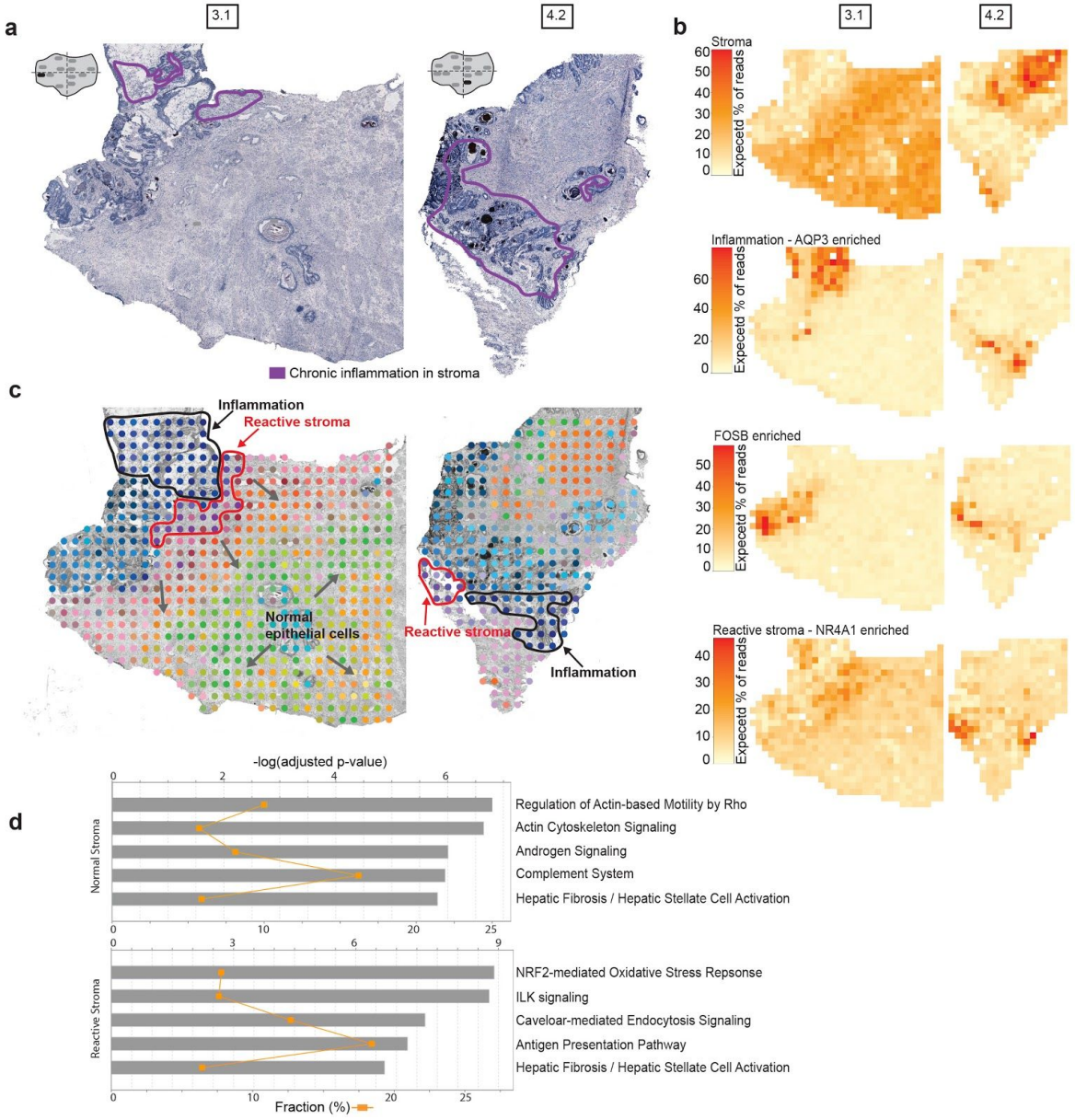
Histology and shared gene expression factors in three cancer tissue sections. **a** Annotated brightfield images of H&E-stained tissue sections. **b** Factor activity maps for selected factors corresponding to normal glands, stromal, PIN, cancerous, or inflamed regions based on a joint factor analysis of the three sections. Remaining factors' activity maps in Supplementary Figure 6 and Supplementary data 2. **c** Pathways enriched in normal, cancer and PIN epithelium clearly differentiate healthy from diseased tissue.

Figure 4


a
b
Genes up-regulated in periphery
Genes up-regulated in center
Cancer area chosen for normalisation
Cancer center spots
Cancer periphery spots
c
ACTB
ASAH1
ACTG2
KLK4
EGFL7
NUPR1
TAGLN
PDLIM5
periphery
center
log10(p-value)
log10(fold change)
gene with p-value < 0.01
gene with p-value < 0.05
gene with p-value > 0.05
d
Pathways enriched in periphery
Pathways enriched in center
Protein processing in endoplasmic reticulum
Oxidative phosphorylation
Lysosome
Huntington's disease
Phagosome
Endocytosis
Pentose phosphate pathway
Citrate cycle (TCA cycle)
Glycolysis / Gluconeogenesis
Regulation of actin cytoskeleton
Focal adhesion
MAPK signaling pathway
ECM−receptor interaction
PI3K−Akt signaling pathway
NF−kappa B signaling pathway
Toll−like receptor signaling pathway
NOD−like receptor signaling pathway
T cell receptor signaling pathway
Fructose and mannose metabolism
Chemokine signaling pathway
Ras signaling pathway
p-value
highly significant
Pathway enrichment
significant
e
f
g
h
Protein processing in endoplasmic reticulum
Lysosome
Endocytosis
Phagosome
Regulation of actin cytoskeleton
Focal adhesion
PI3K−Akt signaling pathway
AMPK signaling pathway
Ras signaling pathway
TNF signaling pathway
Pathways in cancer
Chemokine signaling pathway
ECM−receptor interaction
Prostate cancer
Wnt signaling pathway
MAPK signaling pathway
Fructose and mannose metabolism
Oxidative phosphorylation
Glycolysis / Gluconeogenesis
Huntington's disease
Pyruvate metabolism
Amino sugar and nucleotide sugar metabolism
Pentose phosphate pathway
Citrate cycle (TCA cycle)
HIF−1 signaling pathway

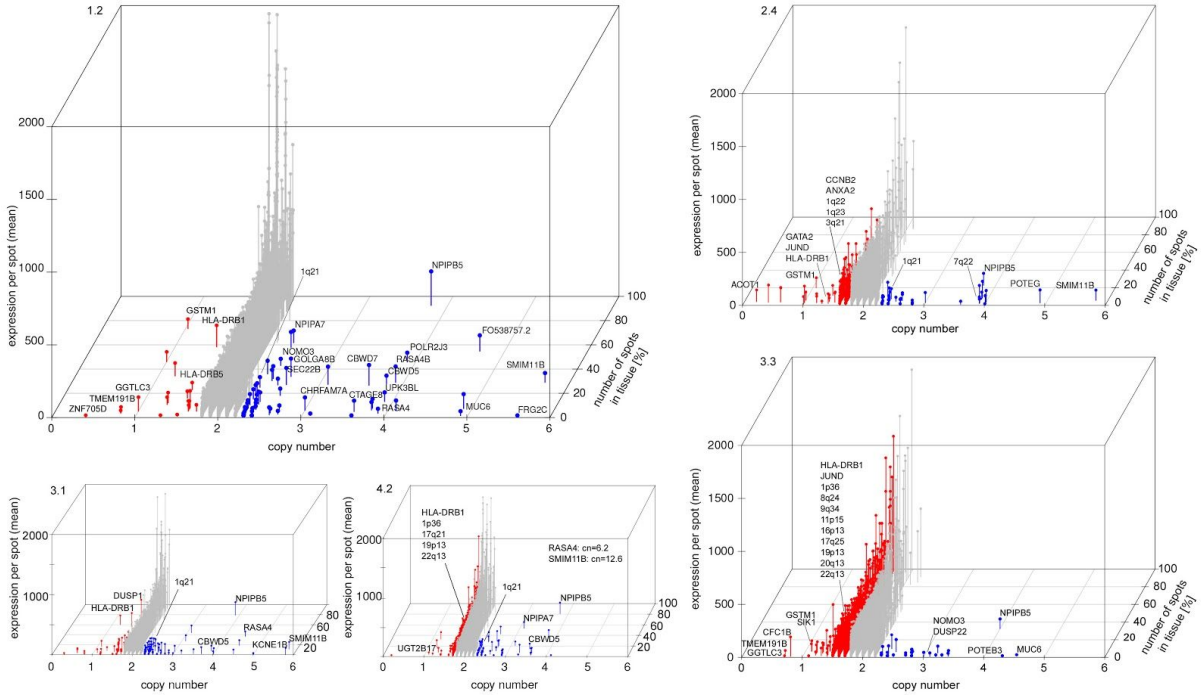
Spatial comparison of periphery and center of tumor transcriptomes. **a-d** Tissue sample 1.2 **e-h** Tissue sample 3.3 **a, e** Area comprising spots taken for normalisation of ST counts, within this area spots are chosen as periphery and center. Choice of spots is based on the pathologist's annotation and the activity of the factors "cancer" and "reactive stroma" **b, f** Volcano plot of significantly differentially expressed genes between periphery and center **c, g** Box plots showing expression levels of noteworthy genes significantly upregulated in either periphery or the cancer center **d, h** Enriched pathways for significantly ($p < 0.05$) differentially expressed genes in center and periphery

Figure 5



Stromal heterogeneity and reactive stroma in the microenvironment of inflammation. **a** Annotated brightfield images of H&E-stained tissue sections. **b** Selected factor activity maps of two inflammation-related factors and two stromal factors (normal and reactive) based on a joint factor analysis of tissue sections 3.1 and 4.2. Remaining factors' activity maps in Supplementary Fig. 22 and Supplementary data 4. **c** t-SNE summary of factor activities of all factors from the analysis in (B); similar colors indicate similar factor activities. Arrows indicate some stromal expression gradients. **d** Top five enriched pathways in reactive and normal stroma. Bars give significance and orange squares the ratio. P-values were corrected for multiple testing by the Benjamini-Hochberg procedure.

Figure 6



Relationship of copy number and gene expression. Expressed genes with a copy number between zero and six are shown. The majority of expressed genes have a copy number of two. Sample 1.2 is characterized rather by amplifications, sample 3.3 rather by deletions, whereas sample 2.4 displays a mixture of deletions and amplifications. Besides the deletions and amplifications that are unique for one or more samples, we identify several germline mutations. For example, we observe a heterozygous deletion of GSTM1 (CN=1) which is linked to "an increased susceptibility to environmental toxins and carcinogens"[48] and a germline deletion at 11q11 (CN=0) which is linked to obesity[49]. Somatic deletions were found among all 12 samples in USP17L8 (CN<0.5)